INSTITUTO SUPERIOR TÉCNICO

PROJETO INTEGRADOR DE 2º CICLO

Applying Emerging Deep Learning Techniques to General Relativity

---

Author:
Guilherme SIMPLÍCIO

Supervisors:
Dr. Carlos HERDEIRO
Dr. Richard BRITO

Research work performed for the Master in Engineering Physics

at

Gr@v
Physics Department UA
&
Centra
Physics Department IST

January 16, 2023

1 Introduction

What happens to a star when it dies [1]? Chandrasekhar in 1931[6], improving the previous work of E. C. Stoner[22] proved that there is a critic limit where white dwarfs are not stable, thus collapsing (ie, their own gravitational self attraction is not balanced by electron degeneracy pressure) and forming Black Holes (BH)[2]. This calculation was initially overlooked by the community of scientists because such a limit would impose the existence of BH, which were considered a scientific impossibility at the time but showed a glimpse of what GR (General Relativity) had for us in the future. With the further improvement of computational techniques at the disposal of Physics (also sometimes overlooked in the Physics community) one can hope that a new generation of GR starts now with more exciting discoveries.

1.1 General Relativity in a nutshell

Newton was able to see further, by standing on the shoulders of giants. Fortunately Einstein stood on the shoulders of Newton and could explain misunderstood phenomena such as the precession of the perihelion of Mercury. With the Principle of Equivalence, Principal of General Covariance and the Einstein's Postulate, Einstein postulated the Einstein's Field Equations (EFE) [10]:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \tag{1}$$

$G_{\mu\nu}$ is the Einstein tensor (related with the Ricci Tensor and Ricci Scalar), $T_{\mu\nu}$ is the energy-momentum tensor and Λ is the cosmological constant, that can be thought as some form of energy with negative pressure. This rather simple looking equation, is a system of second order partial differential equations and relates central quantities, stating that the curvature of the space-time is directly related to the energy and momentum of whatever matter and radiation are present [11]. The solutions of equation (1) are not trivial, Birkhoff's theorem states that any line element which is both a solution to EFE in vacuum ($T_{\mu\nu}$= 0) and spherically symmetric, is static and asymptotically flat, which can be described by the Schwarzschild line element [25]:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{2}$$

with $d\Omega^2 = d\theta^2 + sin^2\theta d\phi^2$. Birkhoff's theorem is quite powerful, it means Schwarzschild solution[3] does not just describe a BH, but it describes the space-time outside any non-rotating, spherically symmetric object, like a star. This solution assumes $\Lambda = 0$, we will briefly return to this point after a quick analysis on (2). This solution seems to diverge when r = 2M, however this is only an apparent singularity, as a simple change to Eddington-Finkelstien coordinates shows [9] [12] [21]. Using this method the line element is analytically expanded for $r > 0$. There is nothing at r = 2M to prevent the star collapsing through r = 2M. Considering now the EFE with $\Lambda > 0$, with the same assumptions as the previous solution, the metric takes the form:

$$ds^2 = -\left(1 - \frac{2M}{r} - \frac{r^2}{R^2}\right) dt^2 + \left(1 - \frac{2M}{r} - \frac{r^2}{R^2}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{3}$$

with $R^2 = \frac{3}{\Lambda}$. (3) is called de Sitter–Schwarzschild spacetime [8]. One could also do the calculations for $\Lambda < 0$, (anti de Sitter space). However, the numerical value of Λ has been calculated in excellent

[1]meaning, once it runs out of fuel (hydrogen) in its core.
[2]Note that similar calculations have also been done for Neutron Stars.
[3]Interestingly enough Schwarzschild found the solution while serving the German army during World War I.

agreement with the measurements obtained by the High-Z Supernova Team and the Supernova Cosmology Project [5]. Still in the domain of spherically symmetric solutions, the solution for charged BH was discovered by Reissner [24] and Nordström [20] independently and it is

$$ds^2 = -\left(1 - \frac{2M}{r} - \frac{Q^2}{r^2}\right)dt^2 + \left(1 - \frac{2M}{r} - \frac{Q^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{4}$$

with Q the charge of the BH. In this (charged) case, Birkhoff's theorem guarantees the line element is representative of the Reissner-Nordström(RN) geometry. All the other solutions previously stated, although insightful, assume that the BH is static. This is generally not true for a given object in the universe. The so called Kerr-Newman solution [16][18][19], in the Boyer-Linquist coordinates is

$$\begin{aligned} ds^2 = &-\frac{(\Delta - a^2\sin^2\theta)}{\Sigma}dt^2 - 2a\sin^2\theta\frac{(r^2+a^2-\Delta)}{\Sigma}dtd\phi \\ &+\left(\frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\right)\sin^2\theta d\phi^2 + \frac{\Sigma}{\Delta}dr^2 + \Sigma d\theta^2 \end{aligned} \tag{5}$$

where

$$\begin{aligned} \Sigma &= r^2 + a^2\cos^2\theta \\ \Delta &= r^2 - 2Mr + a^2 + e^2 \end{aligned}$$

The three parameters are $M, a,$ and e. It can be shown that $a = \frac{J}{M}$ where J is the total angular momentum, while $e = \sqrt{Q^2 + P^2}$ and Q and P are the electric and magnetic (monopole) charges, respectively. Note that setting a = 0, one obtains the RN solution, with magnetic charge. In these conditions, the *No-hair theorem* states that BH are characterised by only three numbers: mass M, angular momentum J and charge Q. Assuming the principle of quasi-neutrality, it is fair to assume that charged black holes do not exist, models considering $e \neq 0$ are still important in the sense that they give a glimpse of the physics in BH [27].

1.2 Gravitational Waves, the ripples of space-time

When we clap our hands, which surely weight no more than 1 kg, a wave is emitted, depending on the force of the collision it can be quite *audible*. Now lets imagine that two BH collide, this violent event emits a (lot of) energy. To put it in perspective, on September 14th, 2015 it was the first time Gravitational Waves (GW) were detected [1] yet the collision of the two BH that originated that event was roughly 1.3 billion years ago.

As previously stated, the EFE are difficult to solve and still have no general solution. However, progress can be made by considering situations in which the metric is almost flat and $\Lambda = 0$. The metric takes the form of

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{6}$$

where $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$ is the Minkowski metric, which implies flat space-time, no curvature. The components $h_{\mu\nu}$ are assumed to be small perturbation of this metric: $h_{\mu\nu} \ll 1$. The usual procedure is to expand the EFE to linear order in the small perturbation $h_{\mu\nu}$. At this order, we can think of gravity as a symmetric "spin 2" field $h_{\mu\nu}$ propagating in flat Minkowski space $\eta_{\mu\nu}$. In Linearised Gravity, as in Electromagnetism, it is useful to impose a gauge. The analogous to the Lorenz gauge, is the *de Donder* gauge

$$\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h = 0 \tag{7}$$

Linearizing (6) and imposing the de Donder gauge(7) in (1), the EFE becomes

$$\Box h_{\mu\nu} - \frac{1}{2}\Box h\eta_{\mu\nu} = -16\pi T_{\mu\nu} \tag{8}$$

With $\Box = \nabla^{\mu}\nabla_{\mu} = g^{\mu\nu}\left(\partial_{\nu}\partial_{\mu} - \Gamma^{\rho}_{\nu\mu}\partial_{\rho}\right)$, the d'Alembertian operator. As we fixed the gauge, the connection term vanishes and we simply have $\Box = g^{\mu\nu}\partial_{\mu}\partial_{\nu}$. Defining $\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}$ and taking the trace of both sides gives $\bar{h} = \eta^{\mu\nu}\bar{h}_{\mu\nu} = -h$ so, given $\bar{h}_{\mu\nu}$ we can trivially reconstruct $h_{\mu\nu}$ as

$$h_{\mu\nu} = \bar{h}_{\mu\nu} - \frac{1}{2}\bar{h}\eta_{\mu\nu} \tag{9}$$

Written in terms of $\bar{h}_{\mu\nu}$, the linearised Einstein equations in de Donder gauge are reduced to a sysyem of wave equation

$$\Box\bar{h}_{\mu\nu} = -16\pi T_{\mu\nu} \tag{10}$$

and we can use what we have learned in electromagnetism[4], to learn something about gravity [26].

1.3 Machine Learning, a black box?

Artifical Intelligence (AI) is defined by the Oxford Dictionary as *the theory and development of computer systems able to perform tasks normally requiring human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages*. Machine Learning (ML) is a subset of AI that focuses on the study of algorithms that learn to make predictions from a dataset without being explicitly programmed to do so. ML has some important landmarks, probably the most famous one occurred in 1997, when DeepMind's Deep Blue, a chess computer, beat Garry Kasparov, chess world champion at the time, under regular time controls [4]. DeepMind has also succeeded, at least partially, in solving one of the biggest unsolved problems in biology: the prediction of the 3D structure of a protein based solely on its amino acids sequence[15]. Even though ML has all this impressive accomplishments, the mathematical principles underlying AI and ML, however, is still not rigorously explained. For many successful applications a through mathematical explanation is still missing, thus the often denotation of *black box*. Is there a way to give physical meaning to this great tool? In Section 2 we give an answer to this question.

1.3.1 Neural Networks

Neural Networks (NN) are a subset of ML and are at the heart of Deep Learning (DL) algorithms. Both their name and algorithm is based upon the human brain, the node layers (input, hidden and output) mimic the way that biological neurons signal to one another. An example of a simple fully connected NN is shown in figure 1.

A forward fully-connected neural network (i.e., each neuron of a certain layer is connected to the all neurons of the successive layer), where we call L the number of layers, $X \in \mathbb{R}^{n\times d}$ the inputs and $y \in \mathbb{R}^{n}$ the labels, with n the number of samples, and d the input dimension. We also define p_l as the number of neurons in the l^{th} layer, with $l = 1, \ldots, L$. When considering regression problems, we have that the size of the output layer, namely p_L is equal to 1 , while we can have $p_L = k$ in the case of classification problems, where k is the number of classes. In a NN , each node, or artificial neuron, connects to another and has an associated weight and threshold. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along to the next layer of the network. Note that NN rely on training to improve accuracy. Each node of the NN is its own linear regression model[5]

Once an input layer is determined, weights are assigned. These weights determine the importance of any given variable, with larger ones contributing more significantly to the output compared to other

[4] The famous Green's Function comes in handy.
[5] Linear Regression is a key concept in ML, for a quick but thorough overview, refer to IBM's website.

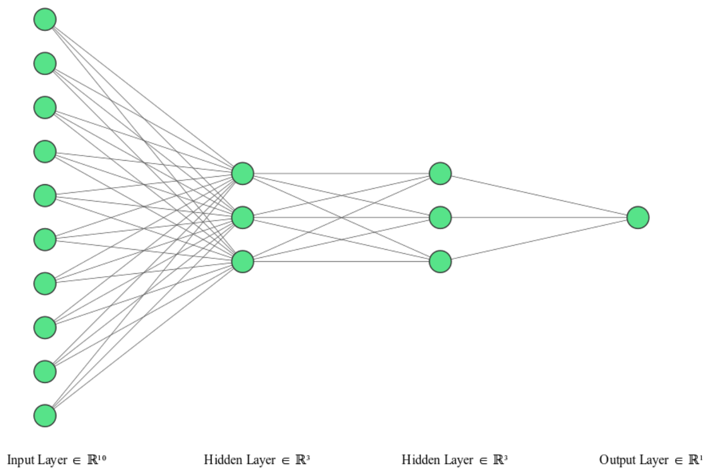


FIGURE 1: 2-hidden layer feed forward fully connected neural network with D = 3, input dimension of 10 and output dimension of 1. It was generated using this website.

inputs. All inputs are then multiplied by their respective weights and then summed. Afterward, the output is passed through an activation function, which determines the output. If that output exceeds a given threshold, it "fires" (or activates) the node, passing data to the next layer in the network. This results in the output of one node becoming in the input of the next node. This process of passing data from one layer to the next layer defines this neural network as a feedforward network [14]. Ultimately, as in ML the objective is to minimize our cost function to ensure correctness of fit for any given observation.

2 State of the art

Machine Learning and AI in general needs to be advanced and understood better in order to unleash its full potential in science. When using machine learning, a physicist should be required not only to apply learned techniques, but also to understand them. This is where Scientific Machine Learning (SciML) comes in. SciML is an emerging branch of ML that gives meaning and interpretability to the *black box*. This branch is fairly new in GR but has been utilized for some years in other fields of physics, such as Geophysics, which heavily depends on the resolution of highly complicated differential equations.

2.1 Physics-informed neural networks [17]

Physics-informed neural networks (PINNs) were defined by Maziar Raissi et al, as "neural networks that are trained to solve supervised learning tasks while respecting any given law of physics described by general nonlinear partial differential equations" [23]. This means that the black box aspect of the ML, is somewhat minimized by encoding the Physics in the Loss Function of the NN, the function that is minimized! The idea of PINNs is therefore to "encode" the laws of physics and scientific knowledge into learning algorithms so as to make them more robust and to improve their performance, this embedding of prior knowledge in the NN limits the space of solutions, and makes learning the algorithm easier, thus being able to generalize better with a lower amount of training data. This physical knowledge can be encoded in the loss function in a number of ways, such as: governing equations, boundary conditions, symmetries and any other creative way one has to limit the solutions of the system.

Schematically speaking a PINN consists in a NN(x,θ) with input x and trainable parameters θ. This PINN is designed to model the solution of a differential equation u(x), by using a NN, hence NN(x,θ)

$\approx$ u(x). In order to train the PINN, a loss function is used, usually composed of at least two terms, a boundary loss $L_b(\theta)$, which tries to restrain the domain of the solutions: matching the PINN solution to the known solution (and/or its derivatives) along the boundaries of the domain. There is also a physics loss $L_p(\theta)$, which tries to minimise the residual of the underlying equation at a set of locations within the domain. Therefore, a typical loss function can be summarized as

$$L(\theta) = \alpha L_p(\theta) + \beta L_b(\theta) \tag{11}$$

Note the importance of multiplying each loss by a factor so each loss has "equal importance" being minimized. Obviously with a sum of just two factors one of the constants can be set to 1, but it is important to state this idea for loss functions with sums of more than two terms. The derivatives of the input, concerning the differential equations, are obtained using autodifferentiation. One could say that training a PINN is basically solving a differential equation. However, PINNs require much less training data and are able to extrapolate away from these example solution points. PINNs also have other advantages over traditional methods of solving differential equations such as finite difference and finite element modelling, for instance, providing a continuous, mesh-free solution. Additionally, they provide analytical and tractable gradients with respect to their inputs. Lastly, PINNs can also be used for inverse problems, in which case an additional “data” loss is added, usually ordinary least squares, which compares the PINN solution with known solution values at additional locations within the domain, and parameters of the underlying equation are jointly optimised alongside θ. Note that this additional "data" loss can be added also in the *prediction* problem, and not only in the *inference* problem, to improve accuracy.

Finally, PINNs are not unfortunately perfect, they struggle to converge as they have millions of free parameters and it is not yet clear how to optimize them. Furthermore, when used simply for forward simulation tasks, PINNs are usually less efficient than traditional approaches such as finite difference and finite element methods, because of their large optimisation costs and the fact they need to be retrained for each simulation. However, PINNs can outperform traditional approaches when used for inversion problems, where traditional inversion algorithms often require thousands of forward simulations to converge.

2.1.1 PINNs applied to the Schrödinger equation [6]

In a first attempt of using PINN’s the time dependent 1D Schrodinger equation was chosen:

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi(x,t)}{dx^2} + (V(x) - E)\Psi(x,t) = 0 \tag{12}$$

By Separation of variables, it has the solution $\Psi_n(x,t) = \psi_n(x) * \phi_n(t)$ with $\phi(t) = e^{-i\frac{E_n * t}{\hbar}}$, we can focus on the spatial coordinate, $\psi_n(x)$. Now, the problem depends on the potential, for simplicity we shall consider the well know infinite square well, of length L:

$$V(x) = \begin{cases} 0 & , 0 \leq x \leq L \\ \infty & \text{, otherwise} \end{cases} \tag{13}$$

With the initial conditions $\psi(0) = 0 = \psi(L)$, has the following exact solution:

$$\psi_n(x) = \sqrt{\frac{2}{L}} sin(\frac{n\pi}{L}x) \tag{14}$$

[6] All of the code developed is freely available on my GitHub.

with n=1,2,3,... Noting that we set $k = \frac{\sqrt{2mE}}{\hbar}$, thus the initial conditions, that have $k_n = \frac{n\pi}{L} \Rightarrow E_n = \frac{n^2\pi^2\hbar^2}{2mL^2}$. Hence, we know the full stationary states:

$$\Psi_n(x,t) = \sqrt{\frac{2}{L}} sin(\frac{n\pi}{L}x)e^{-i\frac{n^2\pi^2\hbar}{2mL^2}t} \tag{15}$$

Figure 2 shows us the analytical solution for L=2 and 10 mere data points, which serve for training.

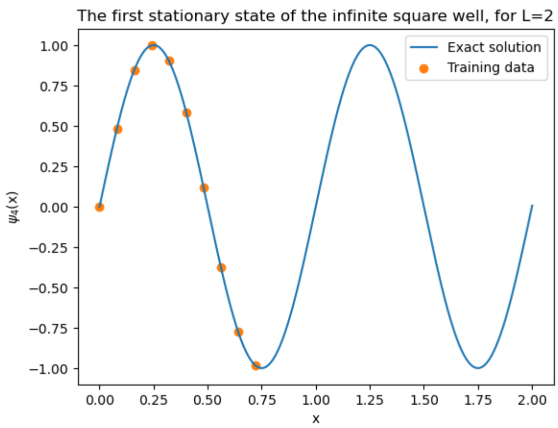
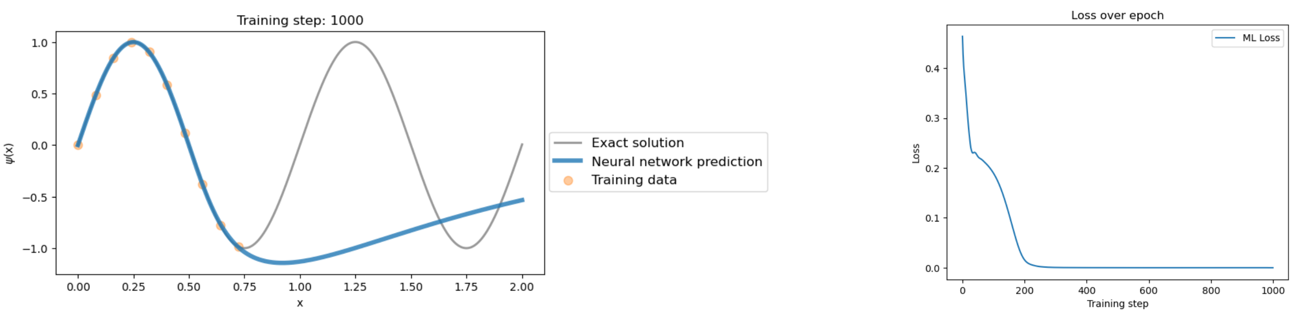


FIGURE 2: First stationary state of the infinite square well for L = 2, n = 4.

With these 10 points, we tried, with different activation functions to predict the solution in the full domain. An Activation Function decides whether a neuron should be activated or not, for our case, GeLU, ReLU and Tanh were the ones used. In short, Tanh is highly popular because of being infinitely differentiable with regards to the input coordinates, ReLU stands for Rectified Linear Unit, it has a derivative function and allows for backpropagation while simultaneously making it computationally efficient and GeLU means Gaussian Error Linear Unit, usually better than ReLU because of its nonlinearity [2]. For now, our loss function is only data driven, the ordinary least squares.



(A) Prediction of the NN.

(B) Loss function over training step.

FIGURE 3: Prediction and training loss plots for the tanh activation.

Some conclusions can be drawn. Not one activation was able to converge, the data points are not sufficient for the NN. Note that the loss actually gets minimized to real low values, the system is doing its best, but with just 10 points it is impossible to converge.

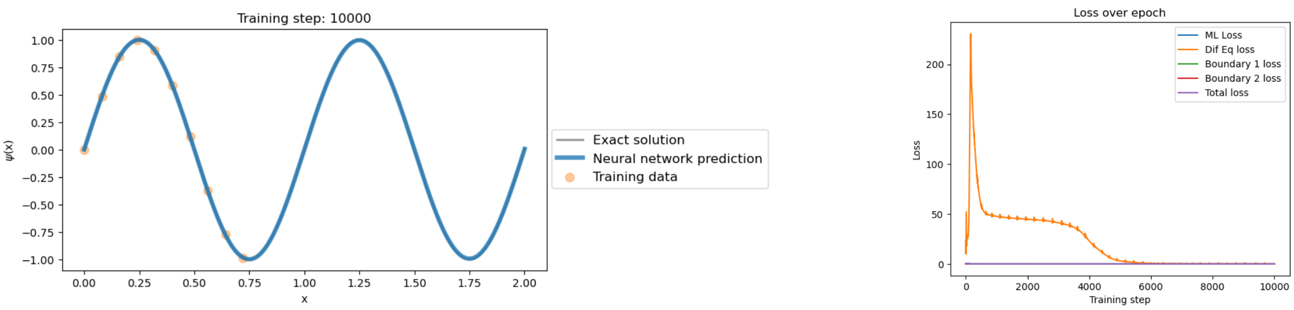
We will define three "Loss functions":

- One that is data driven, typical in ML, that minimizes the distance between the data points and the model of the NN;

- Two that are physics driven:
  - One that takes care of the boundary conditions;
  - One that gives the equation to the network.

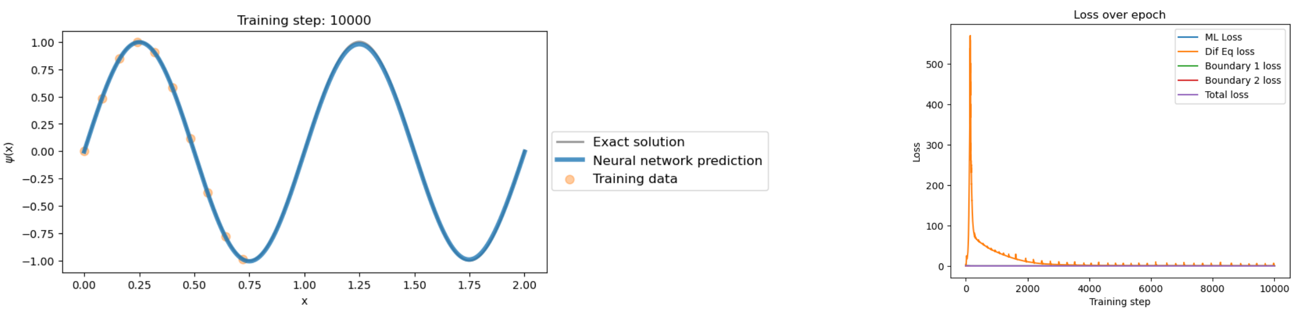
Once again we intend to minimize the loss, ie, force the network to respect the differential equation, in this case, the Schrodinger Equation, the Boundary Conditions and of course, let the black box do what it does best. As noted in (11), one has to take into account the order of magnitude of each "Loss function" and balance it out, meaning:

$$Loss = \lambda_1 Loss_1 + \lambda_2 Loss_2 + \lambda_3 Loss_3 + \lambda_4 Loss_4 \tag{16}$$



(A) Prediction of the NN. (B) Loss function over training step.

FIGURE 4: Prediction and training loss plots for the tanh activation, with PINNs.



(A) Prediction of the NN. (B) Loss function over training step.

FIGURE 5: Prediction and training loss plots for the GeLU activation, with PINNs.

Amazingly, the NN is able to converge to the exact solution, with just 10 data points, as a normal NN could not! With this test, one could also check that the GeLU activation [13] seems to converge faster, hence it is the one used throughout this exercise.

Now, done with this problem, the Bohr potential is tackled. The radial equation for the Bohr potential $V(r) = -\frac{e^2}{4\pi\epsilon_0}\frac{1}{r}$ is:

$$-\frac{\hbar^2}{2m}\frac{d^2u}{dr^2} + \left(-\frac{e^2}{4\pi\epsilon_0}\frac{1}{r} + \frac{\hbar^2}{2m}\frac{l(l+1)}{r^2} - E\right)u(r) = 0 \tag{17}$$

We will skip the details but once again it has a separable solution, $\Psi_{nlm}(r,\theta,\phi) = R_{nl}(r)Y_\ell^m(\theta,\phi)$, $Y_\ell^m(\theta,\phi)$ are the spherical harmonics. Y depends on two integers, l that is called the azimuthal quantum number and m the magnetic quantum number, these are physical quantities that are related with the angular momentum of the electron. We also have n, the principal quantum number: it tells you the energy of the electron. All this will be important briefly. The radial solution is:

$$R_{nl}(r) = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!}} e^{-r/na} \left(\frac{2r}{na}\right)^\ell \left[L_{n-\ell-1}^{2\ell+1}(2r/na)\right] \tag{18}$$

with, $a = 0.52910^{10}m$ (Bohr radius) and $L_q^p(x)$ are the associated Laguerre polynomials. Although not pretty the full solution is:

$$\psi_{n\ell m} = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!}} e^{-r/na} \left(\frac{2r}{na}\right)^\ell \left[L_{n-\ell-1}^{2\ell+1}(2r/na)\right] Y_\ell^m(\theta,\phi) \tag{19}$$

Defining ρ as $\frac{r}{na}$:

$$\psi_{n\ell m} = \sqrt{\left(\frac{2}{na}\right)^3 \frac{(n-\ell-1)!}{2n(n+\ell)!}} e^{-\rho} (2\rho)^\ell \left[L_{n-\ell-1}^{2\ell+1}(2\rho)\right] Y_\ell^m(\theta,\phi) \tag{20}$$

As Griffiths [7] said: don't complain! This is one of the very few realistic systems that can be solved at all, in exact closed form.

Figure 6 shows us the analytical solution for (17), once again with just 10 data points.

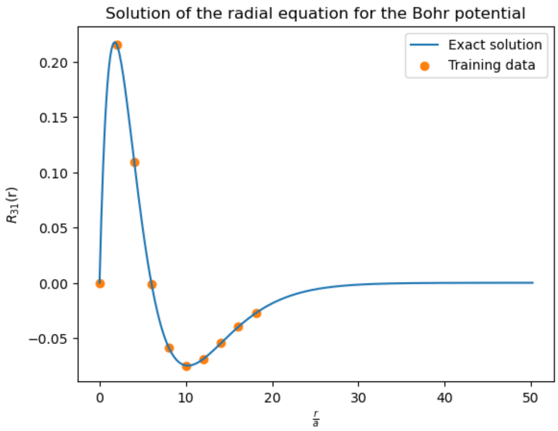
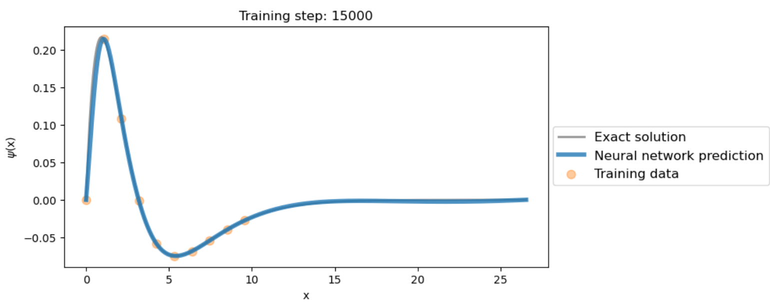
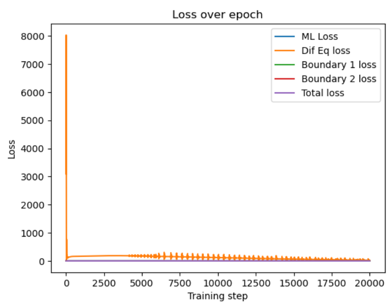


FIGURE 6: Analytical solution for the Bohr potential, with n=3 and l=1.



(A) Prediction of the NN.



(B) Loss function over training step.

FIGURE 7: Prediction and training loss plots of the Bohr Potential problem, for the GeLU activation, with PINNs.

[7] David J. Griffiths has some of the most legendary books in Physics, one of them is the one being referenced, Introduction to Quantum Mechanics.

In figure 7a it is once again noticeable that the system is able to predict the exact solution. In figure 7b the trend that the differential equation loss rules the minimization of the loss function continues. It is the physics that is imposing the accuracy of the system.

As a sanity check, we studied the behaviour of this NN for other quantum numbers, also by varying the position of the training points, ie, the objective was to see if the system would be able to predict the exact solution without having any information of the curve.

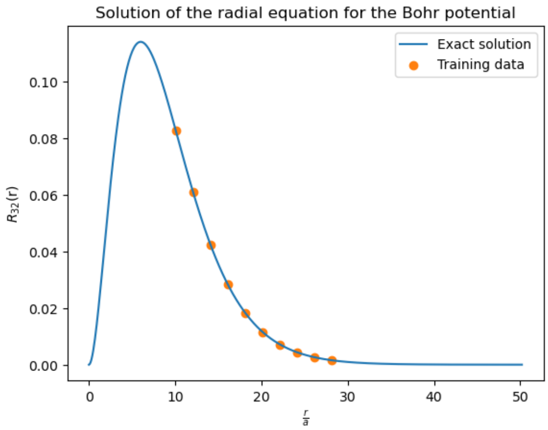
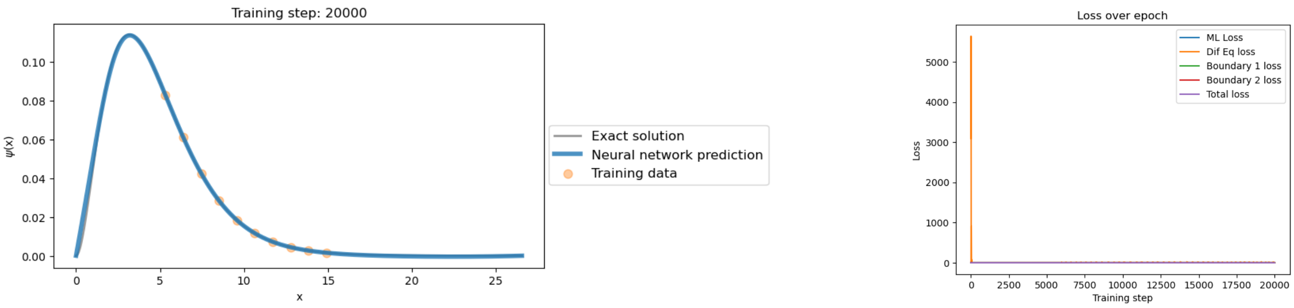


FIGURE 8: Analytical solution for the Bohr potential, with n=3 and l=2.



(A) Prediction of the NN.

(B) Loss function over training step.

FIGURE 9: Prediction and training loss plots of the Bohr Potential problem, for the GeLU activation, with PINNs.

Indeed after some training, the system was also able to predict the exact solution, these results seem quite promising!

The Inversion problem was also studied, the NN not only had the task to predict the behaviour of the function over the domain, but also to estimate a free parameter, in this case, the energy of the system. First it was generated some noisy data with (15), for L = 2, n = 4 (figure 10a). After some training PINNs were able to accurately predict the exact solution. However, it needed some more training to converge to the true value of the energy of the system (figure 11a). It is also important to note that the loss curve (figure 11b) has a slower convergence, prior to what was seen before. Nonetheless, the results were also quite satisfying and encouraging.

(A) Noisy data, Infinite Potential Problem.

(B) Prediction of the NN.

FIGURE 10: Study of Inversion Problem with PINNs

(A) Energy over the training step.

(B) Loss Function over training step.

FIGURE 11: Study of Inversion Problem with PINNs.

3 Objectives & Challenges

Finally, we state the objectives for the thesis. PINNs are a exciting tool and over the next few months a number of interesting topics will be adressed: the study of Q-Balls [7], try to solve the EFE or maybe try to create a library of a continuum spectrum of solutions for the GW. The possibilities are endless and exciting. On the ever growing field of AI, one needs to keep updated on the new developments and the code can surely get better and more efficient, even though if it is quite efficient right now (convergence in a matter of 2 minutes). For instance, JAX[3] is a far quicker extensible system for transforming numerical functions, implementing it would definitely be a major improvement. Also it is important to note that solving the Einstein equation using a PINN is a much more challenging task than solving the Schrödinger equation. This is because the Einstein equation is a much more complex and nonlinear PDE, and it involves many more unknowns and variables. To solve the Einstein equation using a PINN, you would need to first express the equation in a form suitable for numerical solution. Exactly how to do this is still an open question, also for GW.

4 Acknowledgements

Besides the unending support and guidance from my supervisors, Carlos Herdeiro and Richard Brito, I would also like to thank Miguel Zilhão, António Onofre and Osvaldo Freitas for their counselling.

References

[1] B. P. Abbott et al. "Observation of Gravitational Waves from a Binary Black Hole Merger". In: *Phys. Rev. Lett.* 116 (6 2016), p. 061102. DOI: 10.1103/PhysRevLett.116.061102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.116.061102.

[2] Pragati Baheti. *Activation functions in neural networks [12 types amp; use cases]*. URL: https://www.v7labs.com/blog/neural-networks-activation-functions.

[3] James Bradbury et al. *JAX: composable transformations of Python+NumPy programs*. Version 0.3.13. 2018. URL: http://github.com/google/jax.

[4] Murray Campbell, A.Joseph Hoane, and Feng hsiung Hsu. "Deep Blue". In: *Artificial Intelligence* 134.1 (2002), pp. 57–83. ISSN: 0004-3702. DOI: https://doi.org/10.1016/S0004-3702(01)00129-1. URL: https://www.sciencedirect.com/science/article/pii/S0004370201001291.

[5] Moshe Carmeli and Tanya Kuzmenko. "Value of the cosmological constant: Theory versus experiment". In: *AIP Conference Proceedings* 586.1 (2001), pp. 316–318. DOI: 10.1063/1.1419571. eprint: https://aip.scitation.org/doi/pdf/10.1063/1.1419571. URL: https://aip.scitation.org/doi/abs/10.1063/1.1419571.

[6] S. Chandrasekhar. "The Maximum Mass of Ideal White Dwarfs". In: 74 (July 1931), p. 81. DOI: 10.1086/143324.

[7] Sidney Coleman. "Q-balls". In: *Nuclear Physics B* 262.2 (Dec. 1985), pp. 263–283. DOI: 10.1016/0550-3213(85)90286-X.

[8] W. de Sitter. "On the curvature of space". In: 20 I (1918), pp. 229–243. URL: https://www.dwc.knaw.nl/DL/publications/PU00012216.pdf.

[9] A. S. Eddington. "A Comparison of Whitehead's and Einstein's Formulæ". In: 113.2832 (Feb. 1924), p. 192. DOI: 10.1038/113192a0.

[10] Albert Einstein. "Die Feldgleichungen der Gravitation". In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (Jan. 1915), pp. 844–847.

[11] Albert Einstein. "Die Grundlage der allgemeinen Relativitätstheorie". In: *Annalen der Physik* 354.7 (Jan. 1916), pp. 769–822. DOI: 10.1002/andp.19163540702.

[12] David Finkelstein. "Past-Future Asymmetry of the Gravitational Field of a Point Particle". In: *Physical Review* 110.4 (May 1958), pp. 965–967. DOI: 10.1103/PhysRev.110.965.

[13] Dan Hendrycks and Kevin Gimpel. "Bridging Nonlinearities and Stochastic Regularizers with Gaussian Error Linear Units". In: *CoRR* abs/1606.08415 (2016). arXiv: 1606.08415. URL: http://arxiv.org/abs/1606.08415.

[14] IBM. *What are neural networks?* URL: https://www.ibm.com/topics/neural-networks.

[15] Jumper J., Evans R., and Pritzel A. et al. "Highly accurate protein structure prediction with AlphaFold". In: *Nature* 596 (2021), pp. 583–589. DOI: https://doi.org/10.1038/s41586-021-03819-2.

[16] Roy P. Kerr. "Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics". In: 11.5 (Sept. 1963), pp. 237–238. DOI: 10.1103/PhysRevLett.11.237.

[17] B Moseley. "Physics-informed machine learning: from concepts to real-world applications". PhD thesis. University of Oxford, 2022. URL: https://ora.ox.ac.uk/objects/uuid:b790477c-771f-4926-99c6-d2f9d248cb23.

[18] E. T. Newman and A. I. Janis. "Note on the Kerr Spinning-Particle Metric". In: *Journal of Mathematical Physics* 6.6 (June 1965), pp. 915–917. DOI: 10.1063/1.1704350.

[19] E. T. Newman et al. "Metric of a Rotating, Charged Mass". In: *Journal of Mathematical Physics* 6.6 (June 1965), pp. 918–919. DOI: 10.1063/1.1704351.

[20] G. Nordström. "On the Energy of the Gravitation field in Einstein's Theory". In: *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences* 20 (Jan. 1918), pp. 1238–1245.

[21] Roger Penrose. "Gravitational Collapse and Space-Time Singularities". In: 14.3 (Jan. 1965), pp. 57–59. DOI: 10.1103/PhysRevLett.14.57.

[22] Edmund C. Stoner Ph.D. "V. The limiting density in white dwarf stars". In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* 7.41 (1929), pp. 63–70. DOI: 10.1080/14786440108564713. eprint: https://doi.org/10.1080/14786440108564713. URL: https://doi.org/10.1080/14786440108564713.

[23] Maziar Raissi, Paris Perdikaris, and George E. Karniadakis. "Physics Informed Deep Learning (Part I): Data-driven Solutions of Nonlinear Partial Differential Equations". In: *CoRR* abs/1711.10561 (2017). arXiv: 1711.10561. URL: http://arxiv.org/abs/1711.10561.

[24] H. Reissner. "Über die Eigengravitation des elektrischen Feldes nach der Einsteinschen Theorie". In: *Annalen der Physik* 355.9 (Jan. 1916), pp. 106–120. DOI: 10.1002/andp.19163550905.

[25] Karl Schwarzschild. "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie". In: *Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften* (Jan. 1916), pp. 189–196.

[26] David Tong. *David Tong: Lectures on general relativity*. URL: http://www.damtp.cam.ac.uk/user/tong/gr.html.

[27] P. K. Townsend. *Black Holes*. 1997. DOI: 10.48550/ARXIV.GR-QC/9707012. URL: https://arxiv.org/abs/gr-qc/9707012.